Available online at www.sciencedirect.com
SCIENCE DIRECT®
Information and Computation 199 (2005) 87–106
Information and Computation
www.elsevier.com/locate/ic

Canonization for disjoint unions of theories☆

Sava Krstić[a,*], Sylvain Conchon[b]

[a] *Strategic CAD Labs, Intel Corporation, Hillsboro, OR, USA*
[b] *Laboratoire de Recherche en Informatique, Université Paris-Sud, Orsay, France*

Received 1 December 2003; revised 25 June 2004
Available online 28 January 2005

Abstract

If there exist efficient procedures (*canonizers*) for reducing terms of two first-order theories to canonical form, can one use them to construct such a procedure for terms of the disjoint union of the two theories? We prove this is possible whenever the original theories are convex. As an application, we prove that algorithms for solving equations in the two theories (*solvers*) can *not* be combined in a similar fashion. These results are relevant to the widely used Shostak's method for combining decision procedures for theories. They provide the first rigorous answers to the questions about the possibility of directly combining canonizers and solvers.

1. Introduction

In his 1984 paper [19], Shostak proposed a method for combining decision procedures of first-order theories that has influenced the design of several leading tools for automated verification, including PVS [14], SVC [4], and STeP [6]. Shostak's method applies to a collection of signature- disjoint

☆ This work was developed while both authors were affiliated with OGI School of Science & Engineering. The work was funded in part by the NSF Grant CCR-9703218 and a grant from the Intel Corporation.
* Corresponding author. Fax: +1 503 264 4490.
E-mail address: sava.krstic@intel.com (S. Krstić).

0890-5401/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.ic.2004.11.001

theories, where one theory is *free* (entailing valid formulas only) and the others belong to a restricted class, recently christened *Shostak theories*. Each Shostak theory must be convex,[1] and it must have: (1) a *canonizer* that can compute a unique normal form for every term over the theory's signature, and (2) a *solver* that can transform an equation $a \approx b$ between terms into an equisatisfiable set of equations $x_i \approx c_i$ that express the variables x_i occuring in $a \approx b$ as terms c_i over a (possibly empty) set of fresh variables.

Originally, Shostak's method was based on:

Sho-1 An efficient decision procedure for the union of one free theory and one Shostak theory;
Sho-2 The claim that the disjoint union of two (and therefore any finite number of) Shostak theories is a Shostak theory.

It was first discovered in 1996 that there were mistakes in the *Sho-1* algorithm [8]. Finding a correct version of the algorithm became an active research area, and satisfactory solutions have been obtained only recently [16,5,9].

Surprisingly, the validity of *Sho-2* has received minimal serious attention. Shostak himself provided little evidence that this observation was correct. The current status appears to be this:

- Almost all sources restate "the fact" that a canonizer for a disjoint union of theories is easy to obtain from canonizers of individual theories, but no proof is given.
- It is generally accepted that solvers cannot always be combined to produce a solver for the union theory. There are convincing arguments for this, e.g. in [18], but no reasonably complete proof.
- It is also often stated, e.g. in [5], that solvers for some Shostak theories do combine, but without proofs that this happens even for one pair of theories.

This paper is the result of our attempt to understand and prove what can and what cannot be combined. While reasonable definitions for a combination of two canonizers are not difficult to come up with, it is hardly self-evident that the "canonizers" they define satisfy the required properties.We prove in Theorem 4 that combining canonizers indeed goes as expected, assuming that the component theories are convex. The proof requires some effort, and simple counterexamples show that the convexity assumption would be difficult to relax.

In Theorem 5, we prove that under mild assumptions a disjoint union of theories *cannot* have a solver, regardless of the existence of solvers for the original theories. This is a strong negative result, at odds with claims that solvers of some common theories can be combined and at odds with implementations which apparently realize such combinations.

The paper is organized as follows. Section 2 contains preliminary material. The definition of canonizers is given in Section 3. Section 4 defines a candidate canonizer for a combined theory as the normal form function corresponding to a reduction system induced by canonizers of the component theories. The theorem about uniqueness of normal forms is proved in Section 5. Section 6 gives a necessary and sufficient condition for composability of canonizers with an interesting consequence: whether the candidate canonizer for the union theory is indeed a

[1] See definition in Section 2.

canonizer or not is independent of the chosen canonizers of the component theories. Our main results about the (im)possibility of combining canonizers and solvers are given in Sections 7 and 8, respectively.

2. Preliminaries

This section contains a brief survey of adopted (mostly standard) notation.

2.1. Terms

If Σ is a first-order *signature* (a collection of function symbols and relation symbols, with arities), the corresponding set of *terms* will be denoted $T_\Sigma(X)$, where X is some chosen set of variables. Every term is either a variable, a constant (function symbol of arity zero), or of the form $f(t_1,\ldots,t_k)$, where f is a function symbol of arity k and $t_1,\ldots,t_k$ are terms. Terms are standardly visualized as ordered rooted trees whose leaves are labeled with variables and constants, and whose interior nodes are labeled with function symbols of positive arity. Each node has a unique *position* determined by the approach path to it from the root. The position of the root is the empty string ϵ, and if π is the position of some node, then πi is the position of the node's ith child. (For example, in Fig. 1 below, the node labeled "$-$" has position 112.) There is an obvious bijection between positions of a term t and occurrences of subterms of t; the subterm corresponding to the position π will be denoted t_π. (For example, if t is the term depicted in Fig. 1, then $t_{112} = \mathsf{car}(x) - \mathsf{car}(x)$.)

We write $t[\pi \mapsto u]$ for the term obtained by *replacement* of the subterm t_π in t by the term u. Simultaneous replacement of subterms t_π with terms $u(\pi)$, where π belongs to a set P of positions, is denoted $t[\pi \mapsto u(\pi)]^{\pi\in P}$. Note that this is unambiguous only if all positions occurring in P are incomparable (none is a prefix of another).

A *substitution* is any function $\theta: X \to T_\Sigma(X)$ with finite *domain* $\mathsf{dom}(\theta) = \{x \in X \mid \theta(x) \neq x\}$. Its action on terms is a multiple replacement: $\theta(t) = t[\pi \mapsto \theta(t_\pi)]^{t_\pi\in\mathsf{dom}(\theta)}$. A *variable renaming* is an injective substitution whose range is a subset of X.

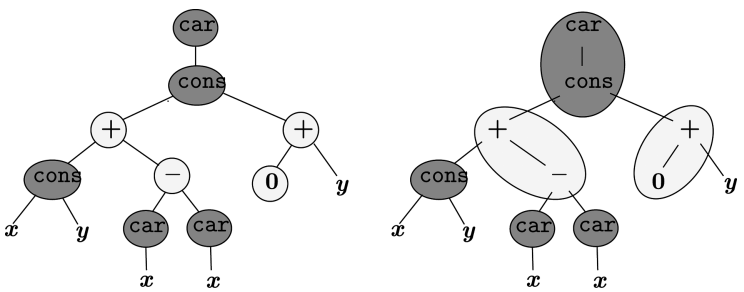

Fig. 1. A term (left) belonging to the disjoint union of arithmetic and the theory of lists, and its blocks (right). Shading indicates different theories.

2.2. Theories

Formulas over Σ are built from atomic formulas using logical connectives $\wedge, \vee, \neg, \longrightarrow, \forall, \exists$. An *atomic formula* is either an *equation* $t \approx t'$, or has the form $p(t_1, \ldots, t_k)$, where p is a relation symbol of arity k, and the t_i's are terms. *Literals* are atomic formulas and their negations. *Disequations* $\neg(t \approx t')$ are written as $t \not\approx t'$.

A Σ-*model* is a non-empty set together with interpretations of symbols in Σ as functions and relations of appropriate arity. In all models, the symbol $\approx$ is interpreted as the equality predicate. Given a Σ-model M, a Σ-formula ϕ, and an assignment ρ of elements of M to free variables in ϕ, we write $M \models_\rho \phi$ if ϕ is true in M under the assignment ρ. A set Γ of formulas is *satisfiable* if $M \models_\rho \Gamma$ (that is, $M \models_\rho \phi$ for every $\phi \in \Gamma$) for some M, ρ. We write $\Gamma \models \phi$ if, for every M and ρ, $M \models_\rho \Gamma$ implies $M \models_\rho \phi$.

A *theory* is a satisfiable set of closed formulas over some signature Σ. If $\mathcal{T}$ and ϕ are a theory and a formula over Σ, we say that ϕ is $\mathcal{T}$-*satisfiable* if $\mathcal{T} \cup \phi$ is satisfiable. Every theory $\mathcal{T}$ defines an equivalence relation on its set of terms: u and v are $\mathcal{T}$-*equivalent* if $\mathcal{T} \models u \approx v$.

A theory $\mathcal{T}$ is called *stably-infinite* if every quantifier-free $\mathcal{T}$-satisfiable formula is true in some infinite model for $\mathcal{T}$.

A theory $\mathcal{T}$ is called *convex* if the validity of the judgment

$$\mathcal{T} \models \psi \longrightarrow u_1 \approx v_1 \vee \cdots \vee u_k \approx v_k$$

where ψ is a conjunction of literals implies that $\mathcal{T} \models \psi \longrightarrow u_i \approx v_i$ holds for some i.

Equational theories, and, more generally, theories closed with respect to the direct product are convex. Note, however, that some important theories (e.g. the theory of arrays) are not convex [13].

2.3. Disjoint unions of theories

Two theories are called *disjoint* if their signatures are disjoint sets.[2] We will use the notation $\mathcal{T}_1 + \mathcal{T}_2$ for the union of disjoint theories. Unions of theories with non-disjoint signatures will not be considered in this paper.

Suppose Σ_1 and Σ_2 are signatures of $\mathcal{T}_1$ and $\mathcal{T}_2$. Define i-*terms* as those terms over $\Sigma_1 + \Sigma_2$ whose root symbol is in Σ_i. Thus, variables are not i-terms for any i. *Pure* i-terms are those whose function symbols are all in Σ_i. The word *mixed* is used for a general (pure or not) term over $\Sigma_1 + \Sigma_2$. *Aliens* of a *mixed* i-term are its maximal non-variable subterms whose top symbol is not in Σ_i. *Alien positions* of t are those π such that t_π is an alien of t. All these definitions obviously extend to unions of more than two signature-disjoint theories.

Mixed terms exhibit a block structure, with blocks corresponding to maximal "pure parts" of the term. Formally, a *block* is a set of positions: two positions π and π' belong to the same block if and only if all symbols occurring on the unique simple path between (and including) π and π' belong to the same Σ_i. An example is given in Fig. 1. Note that alien positions in t are roots of the children blocks of the root block of t. Note also that positions corresponding to occurrences of

[2] The equality symbol $\approx$ is not considered part of theory signatures.

variables are not part of any block, though each such position is clearly associated with a unique block.

3. Canonizers

A *canonizer* for a theory $\mathcal{T}$ would, by the most inclusive definition, be any function $\sigma: T_\Sigma(X) \to T_\Sigma(X)$, which, for a given input u picks a unique representative (the *canonical form*) of the $\mathcal{T}$-equivalence class of u. Thus, a computable canonizer solves the word problem for $\mathcal{T}$. In the literature about Shostak's Algorithm, canonizers are usually required to satisfy also the following properties:

(CAN-1) $\sigma(\sigma(u)) = \sigma(u)$
(CAN-2) $\mathcal{T} \models u \approx v$ if and only if $\sigma(u) = \sigma(v)$
(CAN-3) every variable occurring in $\sigma(u)$ occurs in u
(CAN-4) If $\sigma(u) = u$, then $\sigma(v) = v$ for every subterm v of u

Note that these conditions imply $\mathcal{T} \models \sigma(u) \approx u$. Also, u is a canonical form if and only if $\sigma(u) = u$.

For reasons that will become apparent in Section 4, we will also need to require that canonizers are well-behaved with respect to variable renaming. Full invariance under renaming cannot be expected since, for example, $x + y$ and $y + x$ cannot both be canonical if $+$ is commutative. We will postulate the invariance one normally finds in practice, where preference is defined in terms of an explicit ordering of variables.

Thus, from now on, we will assume a fixed ordering on X that puts the variables in an infinite sequence, and we impose the following condition on canonizers:

(CAN-5) $\sigma(\alpha(u)) = \alpha(\sigma(u))$ for every order-preserving renaming $\alpha: X \to X$ whose domain contains all variables of u

In this paper, a *canonizer* is by definition any function, not necessarily computable, satisfying the five CAN properties. By the following result, the existence of canonizers is guaranteed for all theories with enough ground terms.

Proposition 1. *A canonizer for $\mathcal{T}$ exists if and only if every variable independent term of $\mathcal{T}$ is $\mathcal{T}$-equivalent to a ground term.* (*By definition, t is variable independent if $\mathcal{T} \models t \approx \theta(t)$ for every substitution θ.*)

Proof. Suppose $\mathcal{T}$ has a canonizer and t is a variable independent term that is not $\mathcal{T}$-equivalent to a ground term. Without loss of generality, t is a term in canonical form. Since t contains variables, there exists an order-preserving substitution α such that $t \neq \alpha(t)$. Thus t and $\alpha(t)$ are distinct canonical forms. On the other hand, t and $\alpha(t)$ must be $\mathcal{T}$-equivalent because t is variable independent. This contradiction proves that the existence of canonizers implies that all variable independent terms are equivalent to ground terms.

Turning to the proof in the opposite direction, let us say that a finite set V of variables *supports* a term t if t is $\mathcal{T}$-equivalent to a term that involves only variables from V. Suppose now V and V' both support t and let $W = V \cap V'$. We claim that W also supports t. If $W = \emptyset$, then it is easy to check that t is variable independent, so by assumption W supports t. For the case $W \neq \emptyset$ suppose

$\mathcal{T} \models t \approx t'$, where t and t' contain only variables from V and V', respectively. Then $\mathcal{T} \models t'' \approx t'$, where t'' is obtained from t by substituting variables in $V \setminus V'$ with any other variables. Chosing these other variables from the set W shows that t is supported by W, as claimed.

It follows that for every t there exists a unique smallest set of variables supporting t. Let us call a term *frugal* if it does not contain occurrences of any variables except those belonging to its minimal supporting set.

Let us say now that a set of terms is *transversal* if it

- consists only of frugal terms;
- does not contain two $\mathcal{T}$-equivalent terms;
- is closed under taking subterms;
- is closed under order-preserving variable renamings.

All we need is to show that there exists a transversal set of terms that contains a representative of each class of $\mathcal{T}$-equivalent terms. If S is such a set, then we can define a canonizer σ for $\mathcal{T}$ as the function that maps each term to the unique equivalent term that belongs to S. The properties (CAN 1–5) will clearly be satisfied by σ.

It is easy to see that the family of all transversal sets, ordered by inclusion, is closed under taking unions of chains. By Zorn's Lemma, there exists a maximal transversal set, say S. We claim that S contains a representative of each class of $\mathcal{T}$-equivalent terms. Assume the contrary: there exists a term t such that $t' \notin S$ for any t' that is $\mathcal{T}$-equivalent to t. Without loss of generality, t is frugal and every subterm of t is frugal. Assume also that t is a term with all these properties and minimum possible size. It is easy to see that t is not a variable, so t can be written as $f(t_1, \ldots, t_k)$. The minimality assumption imples that every t_i has a $\mathcal{T}$-equivalent representative $t'_i \in S$. Now $t' = f(t'_1, \ldots, t'_k)$ has no $\mathcal{T}$-equivalent representative in S, while all its subterms are in S. Since t_i and t'_i are both frugal, they contain the same variables, so t' is frugal as well. Let T be the set of all terms $\alpha(t')$, where α is an order-preserving variable renaming. Since S is closed under such renamings, all subterms of terms in T are in S. Thus, $S \cup T$ satisfies the last two conditions for being transversal. It is easy to check that it satisfies the other two conditions as well, so it is a transversal set, contradicting maximality of S. □

Note that the condition for the existence of canonizers given in Proposition 1 is satisfied when the signature of $\mathcal{T}$ contains at least one constant symbol.

4. Combining canonizers

Throughout this section, we assume that $\mathcal{T}_1, \ldots, \mathcal{T}_n$ ($n > 1$) are pairwise disjoint theories with respective signatures $\Sigma_1, \ldots, \Sigma_n$ and canonizers $\sigma_1, \ldots, \sigma_n$. We will write $\mathcal{T}$ for the union theory $\mathcal{T}_1 + \cdots + \mathcal{T}_n$, and Σ for its signature $\Sigma_1 + \cdots + \Sigma_n$. Our goal is to define a function

$$\sigma_1 * \cdots * \sigma_n : T_\Sigma(X) \to T_\Sigma(X)$$

which is a natural candidate for a canonizer of $\mathcal{T}$. It will be obtained as the normal form function of a certain reduction system that canonizers $\sigma_1, \ldots, \sigma_n$ induce on the set $T_\Sigma(X)$ of mixed terms.

4.1. Extending σ_i to Mixed Terms

If t is a (not necessarily pure) i-term, we can still apply the canonizer σ_i to it by treating its alien subterms as variables. For example, the term $cons(x, y) + (car(x) - car(x))$ becomes the pure term $u + (v - v)$ after replacing its alien subterms $cons(x, y)$ and $car(x)$ with fresh variables u, v; the canonizer for linear arithmetic simplifies the pure term into u, so the original mixed term is "canonized" into $cons(x, y)$. To make this extension of σ_i well defined we need to resolve the ambiguities presented by terms like $car(x) + car(y)$, where the result of canonization could depend on the choice of variables used to denote the alien subterms. So let us assume a fixed total ordering of Σ-terms (e.g., lexicographical). Then, given an i-term t, a partial function $\gamma: T_\Sigma(X) \to X$ will be called an *alien abstraction function for* t if

- γ is monotonic (with respect to the given ordering of Σ-terms) and injective;
- the domain of γ contains all alien subterms of t;
- the image of γ does not contain any variable occurring in t.

When γ is an alien abstraction function for t, we write $t \star \gamma$ for the term $t[\pi \mapsto \gamma(t_\pi)]^{\pi \in P}$, where P is the set of all alien positions of t. Thus, $t \star \gamma$ is obtained by replacing the aliens of t with variables specified by γ. We denote by γ^{-1} the obvious substitution $X \longrightarrow T_\Sigma(X)$ that inverts γ.

Definition 1. The *extended canonizer* $\hat{\sigma}_i: T_\Sigma(X) \to T_\Sigma(X)$ is given by

$$\hat{\sigma}_i(t) = \begin{cases} \gamma^{-1}(\sigma_i(t \star \gamma)) & \text{if } t \text{ is an } i\text{-term} \\ t & \text{otherwise} \end{cases}$$

where γ is an alien abstraction function for t.

This definition is a slight modification of the one given by Rueß and Shankar [16,18]; see also [5,11]. Using the propery (CAN-5), it is easy to check that the definition is correct, i.e. independent of the choice of γ.

It follows from Definition 1 that if t is an i-term, then $\hat{\sigma}_i(t)$ is also an i-term, unless, as in our introductory example, $\sigma_i(t \star \gamma)$ is a variable. In such cases, from (CAN-3) we can conclude that $\hat{\sigma}_i(t)$ is either an alien subterm of t or a variable occurring in t . Note also that $\hat{\sigma}_i$ is *not* a canonizer for $\mathcal{T}$.

4.2. Reduction systems for mixed terms

The extended canonizers $\hat{\sigma}_i$ lead immediately to a reduction system $\to$ on the set $T_\Sigma(X)$ of mixed terms. For convenience, we will also consider two smaller reduction systems $\to_I$ and $\to_B$, all defined as follows.

Definition 2. Suppose π is a position in a Σ-term t and suppose the top symbol of t_π is in Σ_i.

(a) If $\hat{\sigma}_i(t_\pi) \neq t_\pi$, we say that π is a *redex* of t and that t *reduces to* $t' = t[\pi \mapsto \hat{\sigma}_i(t_\pi)]$, symbolically $t \to t'$.

(b) We say that π is a *block redex* of t if it is a redex and also the root position of a block of t. The corresponding reduction will be written $t \to_B t'$.
(c) We say that π is an *innermost redex* if it is a block redex and not a prefix of another block redex. Reduction at innermost positions will be denoted $t \to_I t'$.

Example. By inspection, the term t in Fig. 1 has four redexes: ϵ, 11, 12, and 112. The first three are block redexes, the fourth is not. Thus, the innermost redexes are 11 and 12.

Lemma 1. *The reduction systems* $\to, \to_B, \to_I$ *have the same notion of irreducible terms.*

Proof. A position π is a redex of t if and only if the alien abstraction $t_\pi \star \gamma$ is not a canonical form in the corresponding theory $\mathcal{T}_i$. If π is a redex and π' the position of the root of the block containing π, then $t_{\pi'} \star \gamma$ contains $t_\pi \star \gamma$ as a subterm, and by (CAN-4), it too must be a redex. Thus, existence of a redex implies existence of a block redex. Clearly, existence of a block redex implies existence of an innermost redex. □

The following theorem together with Lemma 1 implies the equality of the equivalence relations $\leftrightarrow^*, \leftrightarrow_B^*, \leftrightarrow_I^*$ generated by our three reduction systems.

Theorem 1. *Every equivalence class of* $\leftrightarrow^*$ *contains exactly one irreducible term.*

The obvious approach to proving Theorem 1 by demonstrating local confluence and termination of $\to$ does not work because, as the following example shows, termination is not guaranteed in general.

Example. Let $\mathcal{T}_1$ be the equational theory with one binary symbol f axiomatized by $f(x, y) = f(x, x)$. Let $\mathcal{T}_2$ be any theory with a term u which canonizes to a different term v. It is not difficult to see that there exists a canonizer for $\mathcal{T}_1$ which canonizes $f(x, y)$ to $f(x, x)$, for any variables x, y. Then we have a cyclic derivation: $f(u, v) \to f(u, u) \to f(u, v) \to \cdots$, where in the first step the reduction occurs at the root position of $f(u, v)$, and in the second step it occurs at the root of the second occurrence of u in $f(u, u)$.

The proof of Theorem 1 is rather lenghty and the entire Section 5 is devoted to it. Here we prove a weaker statement that is still sufficient to derive our main results in Sections 7 and 8.

Lemma 2. *Every equivalence class of* $\leftrightarrow_I^*$ *contains exactly one irreducible term.*

Proof. Since the reduction relation $\to_I$ is clearly terminating, it suffices to check that it satisfies the diamond property [1]. Indeed, if the reductions $t \to_I u$ and $t \to_I u'$ correspond to innermost redexes π and π' of t, then π and π' are innermost redexes of u' and u respectively, and reducing u' at π produces the same result as reducing u at π'. □

4.3. The Candidate Canonizer

Theorem 1 shows that there is essentially only one generic way of using the canonizers of the component theories to fully reduce mixed terms. For convenience (to minimize dependence on Theorem 1), we use the innermost reduction strategy in the following definition.

Definition 3. The *candidate canonizer* for $\mathcal{T}$ induced by canonizers $\sigma_1, \ldots, \sigma_n$ is the function $\sigma_1 * \cdots * \sigma_n$ that maps every $\mathcal{T}$-term t to its *normal form*—the unique irreducible term in the $\leftrightarrow_I^*$-equivalence class of t.

Remark. It can be easily proved that the candidate canonizer $\sigma = \sigma_1 * \cdots * \sigma_n$ satisfies

$$\sigma(x) = x$$
$$\sigma(f(t_1, \ldots, t_k)) = \hat{\sigma}_i(f(\sigma(t_1), \ldots, \sigma(t_k)))$$

where x is any variable, and f is any function symbol (in Σ_i, of arity k). These properties are used as a recursive definition for the combined canonizer in [16,18].

It is easy to check that $\sigma_1 * \cdots * \sigma_n$ satisfies all the defining properties of canonizers, except perhaps (CAN-2). We show now that it also always satisfies the soundness part of (CAN-2).

Lemma 3. *Denote* $\sigma = \sigma_1 * \cdots * \sigma_n$. *Then*:

(*a*) $\mathcal{T} \models u \approx \sigma(u)$;
(*b*) *If* $\sigma(u) = \sigma(v)$, *then* $\mathcal{T} \models u \approx v$.

Proof. Part (b) expresses the soundness of our candidate canonizer σ and it follows immediately from Part (a). As for (a), it suffices to prove

$$\mathcal{T} \models u \approx u[\pi \mapsto \hat{\sigma}_i(u_\pi)],$$

where π is the root position of a Σ_i-block of u. Since $u = u[\pi \mapsto u_\pi]$, we only need to prove

$$\mathcal{T} \models \hat{\sigma}_i(u_\pi) \approx u_\pi.$$

With a suitable variable abstraction function γ, we have

$$\hat{\sigma}_i(u_\pi) = \gamma^{-1}(\sigma_i(u_\pi \star \gamma)) \quad \text{and} \quad u_\pi = \gamma^{-1}(u_\pi \star \gamma).$$

Since σ_i is a canonizer, we also have

$$\mathcal{T}_i \models \sigma_i(u_\pi \star \gamma) \approx u_\pi \star \gamma.$$

Combining the last three relations finishes the proof. □

Corollary 4. $\sigma_1 * \cdots * \sigma_n$ *is a canonizer if and only if* $u \not\approx v$ *is* $\mathcal{T}$*-satisfiable for any two distinct irreducible terms* u, v.

Proof. In view of Lemma 3(b) and the remark preceding it, we only need to check the completeness part of (CAN-2), i.e., that $\sigma(u) = \sigma(v)$ must hold whenever $\mathcal{T} \models u \approx v$. By Lemma 3(a), this goal is equivalent to proving that $u \not\approx v$ is $\mathcal{T}$-satisfiable for every two distinct irreducibles u and v. □

In Section 7 we will show that the condition in Corollary 4 is satisfied when the component theories are convex. There, we will also give examples of candidate canonizers that fail to be canonizers.

4.4. Complexity of Combined Canonizers

Clearly, if the canonizers $\sigma_1, \ldots, \sigma_n$ are computable, then the candidate canonizer $\sigma = \sigma_1 * \cdots * \sigma_n$ is computable too. We can sharpen this observation as follows.

Proposition 4. *Suppose $k \geqslant 1$ and each of the canonizers $\sigma_1, \ldots, \sigma_n$ is implemented with time complexity $O(N^k)$, where N denotes the size of the input term. Then the time complexity of any implementation of $\sigma_1 * \cdots * \sigma_n$ that uses innermost reduction strategy is also $O(N^k)$.*

Proof. Assume that the size of trees is measured by the number of nodes and suppose that each of the canonizers σ_i takes time at most cN^k for any input term of size N. Let $P(u)$ denote the set of all root block positions of a Σ-term u that occur as prefixes of innermost redexes of u. We need this easily checked fact about innermost reduction:

> If $u \to_I v$ and $\pi \in P(v)$, then $\pi \in P(u)$, and the blocks of u and v occurring at the position π have equal sizes. Moreover, if the reduction $u \to_I v$ takes place at π, then π is not a redex in v.

Suppose now t is a Σ-term of size N, m is the number of blocks in t, and N_i is the number of nodes in the ith block. It follows from the fact above that bringing t to its normal form can take at most m steps, each associated with a unique block of t. The reduction at any step is an application of an operator $\hat{\sigma}_i$, the essential part of which is a call to σ_i with an input term of size equal to the size of the currently processed block. The total time needed to execute all these calls to canonizers σ_i is thus at most $cN_1^k + \cdots + cN_m^k$. This upper bound is not greater than $c(N_1 + \cdots + N_m)^k = cN^k$. Since $k \geqslant 1$ and the time needed for the rest of the algorithm is clearly $O(N)$, this finishes the proof. □

5. Proof of Theorem 1

Let σ be the candidate canonizer as in Definition 3. Since for every term t, the term $\sigma(t)$ is irreducible and belongs to the equivalence class of t, it suffices to prove that $\sigma(t)$ is the only irreducible in that class. Thus, Theorem 1 can be restated as follows:

$$\sigma(t) = \sigma(t') \text{ holds for every } t, t' \text{ such that } t \to t'. \tag{1}$$

We start with three lemmas.

Lemma 5.

(*a*) *If u is a Σ-term and ρ is the root position of a block of u, then $u \to_I^* u[\rho \mapsto \sigma(u_\rho)]$.*
(*b*) *Suppose θ is a substitution such that, for every x in its domain, $\theta(x)$ is not an i-term. Then $\theta(u) \to_I^* u(\sigma \circ \theta)$, for every pure i-term u.*

Proof. (a) If ρ is a root block position in u and $u_\rho \to_I v$, then clearly $u \to_I u[\rho \mapsto v]$. Consequently, if ρ is a root block position in u, and $u_\rho \to_I^* v$, then $u \to_I^* u[\rho \mapsto v]$. The statement of the lemma follows from this by taking $v = \sigma(u_\rho)$.

(b) Let P be the set of all positions π in u such that u_π is a variable belonging to the domain of θ. By assumption, every $\pi \in P$ is an alien position in $\theta(u)$. Thus, we have

$$\begin{aligned}\theta(u) &= u[\pi \mapsto \theta(u_\pi)]^{\pi\in P} = \theta(u)[\pi \mapsto \theta(u_\pi)]^{\pi\in P}\\ &\to_I^* \theta(u)[\pi \mapsto \sigma(\theta(u_\pi))]^{\pi\in P} = u[\pi \mapsto \sigma(\theta(u_\pi))]^{\pi\in P} = \sigma \circ \theta(u)\end{aligned}$$

where the middle step is justified by part (a) of the lemma. □

Lemma 6. *Suppose u is a pure i-term, $\theta: X \to T_\Sigma(X)$ is a substitution, and $\gamma: T_\Sigma(X) \to X$ is an alien abstraction function for $\theta(u)$. Then*

$$\theta(u) \star \gamma = \bar{\theta}(u),$$

for some substitution $\bar{\theta}: X \to T_{\Sigma_i}(X)$ that depends only on θ and γ.

Proof. For a precise description of alien positions and alien subterms of $\theta(u)$ we need to partition $\mathsf{dom}(\theta)$ into three subsets X_1, X_2, X_3 defined by

$$\begin{aligned}x \in X_1 &\quad \text{iff} \quad \theta(x) \text{ is a } j\text{-term for } j \neq i\\ x \in X_2 &\quad \text{iff} \quad \theta(x) \text{ is an } i\text{-term}\\ x \in X_3 &\quad \text{iff} \quad \theta(x) \text{ is a variable}\end{aligned}$$

Alien positions of $\theta(u)$ are either of the form π, where $u_\pi \in X_1$, or of the form $\pi\pi'$, where $\pi \in X_2$ and π' is an alien position in $\theta(u_\pi)$. In the first case, the alien $\theta(u)_\pi$ is just $\theta(u_\pi)$. In the second case, the alien $\theta(u)_{\pi\pi'}$ is $\theta(u_\pi)_{\pi'}$. It is now easy to see that the substitution

$$\bar{\theta}(x) = \begin{cases}\gamma(\theta(x)) & \text{if } x \in X_1\\ \theta(x) \star \gamma & \text{if } x \in X_2\\ \theta(x) & \text{if } x \in X_3\end{cases}$$

satisfies the requirement $\theta(u) \star \gamma = \bar{\theta}(u)$. □

Lemma 7. *Suppose u, v are pure i-terms, $\mathcal{T}_i \models u \approx v$, and θ is a substitution such that $\theta(x)$ is irreducible for every $x \in \mathsf{dom}(\theta)$. Then $\sigma(\theta(u)) = \sigma(\theta(v))$.*

Proof. For both $\theta(u)$ and $\theta(v)$, the root position is the only possible innermost redex. Thus, we only need to prove

$$\hat{\sigma}_i(\theta(u)) = \hat{\sigma}_i(\theta(v)).$$

Let γ be an alien abstraction function for both $\theta(u)$ and $\theta(v)$. In view of Lemma 6, we have $\hat{\sigma}_i(u\theta) = \gamma^{-1}(\sigma_i(\theta(u) \star \gamma)) = \gamma^{-1}(\sigma_i(\bar{\theta}(u)))$, and similarly $\hat{\sigma}_i(\theta(v)) = \gamma^{-1}(\sigma_i(\bar{\theta}(v)))$. This reduces our goal to proving $\sigma_i(\bar{\theta}(u)) = \sigma_i(\bar{\theta}(v))$, which is indeed true, since σ_i is a canonizer, and $\mathcal{T}_i \models \bar{\theta}(u) \approx \bar{\theta}(v)$ is true as a consequence of $\mathcal{T}_i \models u \approx v$. □

We turn now to the proof of the relation (1). The reduction $t \to t'$ happens at some position π, so we have $t' = t[\pi \mapsto \hat{\sigma}_i(t_\pi)]$, for the appropriate i. Let ρ be the root position in t of the block containing π. First we check that it is no loss of generality to assume here that ρ is the root position of t.

Clearly, $t' = t[\rho \mapsto t'_\rho]$, so in view of Lemma 5(a) we have

$$t \to_I^* t[\rho \mapsto \sigma(t_\rho)] \quad \text{and} \quad t' \to_I^* t[\rho \mapsto \sigma(t'_\rho)].$$

For our goal $\sigma(t) = \sigma(t')$, it suffices to prove that $\sigma(t_\rho) = \sigma(t'_\rho)$. Since $t_\rho \to t'_\rho$ (with the reduction taking place at the position π' such that $\pi = \rho\pi'$), this is just the restatement of the original goal with t and t' in place of t_ρ and t'_ρ respectively.

Thus, we can assume $\rho = \epsilon$, so that π is a position within the top block of t. Let a be the pure i-term $t \star \gamma$, where γ is an alien abstraction function for t. Now we have $t = \gamma^{-1}(a)$, $a_\pi = t_\pi \star \gamma$, and $t_\pi = \gamma^{-1}(a_\pi)$. Thus, $\hat{\sigma}_i(t_\pi) = \gamma^{-1}(\sigma_i(a_\pi))$, and from $t' = t[\pi \mapsto \hat{\sigma}_i(t_\pi)]$ we can derive

$$t' = \gamma^{-1}(a)[\pi \mapsto (\sigma_i(a_\pi))\gamma^{-1}] = \gamma^{-1}(a[\pi \mapsto \sigma_i(a_\pi)]),$$

where the second equality is is an instance of the simple fact $\theta(c[\pi \mapsto d]) = \theta(c)[\pi \mapsto \theta(d)]$ that holds for all terms c, d, substitutions θ, and positions π in c.

Using Lemma 5(b), it follows that

$$t' \to_I^* (\sigma \circ \gamma^{-1})(a[\pi \mapsto \sigma_i(a_\pi)]),$$

and also (since $t = \gamma^{-1}(a)$)

$$t \to_I^* (\sigma \circ \gamma^{-1})(a).$$

Now $(\sigma \circ \gamma^{-1})(x)$ is irreducible for every variable x, so Lemma 7 will finally imply $\sigma(t') = \sigma(t)$ as soon as we check that $\mathcal{T}_i \models a[\pi \mapsto \sigma_i(a_\pi)] \approx a$. This is indeed true, because $a = a[\pi \mapsto a_\pi]$ (trivially) and $\mathcal{T}_i \models \sigma_i(a_\pi) \approx a_\pi$ (since σ_i is a canonizer). □

6. Composability of theories

Consider this composability property of a set of theories $\{\mathcal{T}_1, \ldots, \mathcal{T}_n\}$: an equation between mixed terms holds in the union theory $\mathcal{T} = \mathcal{T}_1 + \cdots + \mathcal{T}_n$ if and only if it can be derived by pure equational reasoning from equations that are true modulo some $\mathcal{T}_i$. Theorem 2 below states that this property is a necessary and sufficient condition for the candidate canonizer $\sigma_1 * \cdots * \sigma_n$ to be a canonizer. The theorem has an interesting consequence: the candidate canonizers obtained from various choices of canonizers $\sigma_1, \ldots, \sigma_n$ for $\mathcal{T}_1, \ldots, \mathcal{T}_n$ are either all canonizers or none of them is.

Let us denote by $\equiv_\sigma$ the equivalence relation on Σ-terms induced by the candidate canonizer $\sigma = \sigma_1 * \cdots * \sigma_n$:

$$u \equiv_\sigma v \quad \text{if and only if} \quad \sigma(u) = \sigma(v).$$

Lemma 8.

(*a*) *$a \equiv_\sigma b$ holds for all pure i-terms a, b such that $\mathcal{T}_i \models a \approx b$;*
(*b*) *$\equiv_\sigma$ is a congruence;*
(*c*) *$\equiv_\sigma$ is closed under substitutions; i.e. $u \equiv_\sigma v$ implies $u\theta \equiv_\sigma v\theta$.*

Proof. (a) For pure i-terms, $a \equiv_\sigma b$ holds if and only if $a \equiv_{\sigma_i} b$.

(b) Clearly, $f(t_1,\ldots,t_i,\ldots,t_k) \to f(t_1,\ldots,t_i',\ldots,t_k)$ holds whenever $t_i \to t_i'$ does. Thus, $f(t_1,\ldots,t_k) \to^* f(\sigma(t_1),\ldots,\sigma(t_k))$ holds in general. It follows that the relations $t_1 \equiv_\sigma t_1',\ldots,t_k \equiv_\sigma t_k'$ always imply $f(t_1,\ldots,t_k) \leftrightarrow^* f(t_1',\ldots,t_k')$. Since by Theorem 1 $\equiv_\sigma$ and $\leftrightarrow^*$ coincide, we can conclude that $\equiv_\sigma$ is a congruence.

(c) It suffices to prove that $u \to v$ implies $\theta(u) \equiv_\sigma \theta(v)$. We have $v = u[\pi \mapsto \hat{\sigma}_i(u_\pi)]$ for some π and i, and so $\theta(v) = \theta(u)[\pi \mapsto \hat{\sigma}_i(u_\pi)\theta]$. Since $u = u[\pi \mapsto u_\pi]$, we also have $\theta(u) = \theta(u)[\pi \mapsto u_\pi\theta]$. By Part (b) of the lemma, to show that $\theta(u) \equiv_\sigma \theta(v)$, we only need to prove $\theta(u_\pi) \equiv_\sigma \theta(\hat{\sigma}_i(u_\pi))$.

Let $u = u_\pi \star \gamma$ be the result of the variable abstraction of u_π. We have $u_\pi = \gamma^{-1}(u)$ and $\hat{\sigma}_i(u_\pi) = \gamma^{-1}(\sigma_i(u))$. With this notation, our current goal can be rewritten as

$$(\theta \circ \gamma^{-1})(u) \equiv_\sigma (\theta \circ \gamma^{-1})(\sigma_i(u)) \tag{2}$$

By idempotence of σ, the relation $\theta(t) \equiv_\sigma (\sigma \circ \theta)(t)$ holds for every t and θ. Instantiating this to both sides of (2) transforms (2) into an equivalent form $(\sigma \circ \theta \circ \gamma^{-1})(u) \equiv_\sigma (\sigma \circ \theta \circ \gamma^{-1})(\sigma_i(u))$, which is true as an instance of Lemma 7. □

Theorem 2. *Let $\mathcal{E}$ be the equational theory axiomatized by equations $u \approx v$, where u and v are $\mathcal{T}_i$-equivalent terms for some i. The candidate canonizer $\sigma_1 * \cdots * \sigma_n$ is a canonizer if and only if all $\mathcal{T}$-equivalent terms are $\mathcal{E}$-equivalent.*

Proof. The condition (CAN-2), necessary and sufficient for σ to be a canonizer, can be expressed as the equality of equivalence relations $\equiv_\sigma$ and $\equiv_\mathcal{T}$, the latter being the $\mathcal{T}$-equivalence of terms. Thus, to prove Theorem 2, it remains to check that the equivalence relations $\equiv_\sigma$ and $\equiv_\mathcal{E}$ are the same.

By Birkhoff's Theorem [1], the relation $\equiv_\mathcal{E}$ is the smallest congruence that is closed under substitutions and contains all pure equations in $\mathcal{T}_1,\ldots,\mathcal{T}_n$. In other words, $\equiv_\mathcal{E}$ is the smallest relation satisfying the properties (a)–(c) of Lemma 8. Since $\equiv_\sigma$ satisfies these properties, we have that $\equiv_\mathcal{E}$ is included in $\equiv_\sigma$.

For the opposite direction we need to prove that $u \equiv_\sigma v$ implies $u \equiv_\mathcal{E} v$. This would follow immediately if we can prove that $\sigma(t) \equiv_\mathcal{E} t$ holds for every t. This last goal reduces to proving $t \equiv_\mathcal{E} t'$ under the assumption $t \to t'$. Now $t' = t[\pi \mapsto \hat{\sigma}_i(t_\pi)]$ for some π and i, and our goal becomes $t_\pi \equiv_\mathcal{E} \hat{\sigma}_i(t_\pi)$. If $a = t_\pi \star \gamma$ is the result of the variable abstraction of t_π, we have $t_\pi = \gamma^{-1}(a)$ and $\hat{\sigma}_i(t_\pi) = \gamma^{-1}(\sigma_i(a))$. Thus, it suffices to prove $a \equiv_\mathcal{E} \sigma_i(a)$ which is true because σ_i is a canonizer and so a and $\sigma_i(a)$ are $\mathcal{T}_i$-equivalent. □

Remark. If $\mathcal{T}_1,\ldots,\mathcal{T}_n$ are equational theories, then the equality of $\equiv_\sigma$ and $\equiv_\mathcal{E}$ is clearly satisfied. Thus, Theorem 2 implies that canonizers of equational theories combine. Since equational theories are convex, this fact is also a special case of Theorem 4 below.

7. Convexity and canonization

Recall from Corollary 4 that the existence of two distinct irreducible mixed terms u, v such that $u \approx v$ holds in the union theory is a necessary and sufficient condition for the failure of a candidate canonizer to be a canonizer. This can happen indeed, and for a simple concrete example take the

theory $\mathcal{T}$ with signature consisting of three constants p, q, r constrained by the axiom $p \approx q \vee p \approx r$, and take $\mathcal{T}'$ with one ternary function symbol f constrained by axioms $f(x,x,y) \approx f(x,x,x)$ and $f(x,y,x) \approx f(x,x,x)$. Then $f(p,q,r) \approx f(p,p,p)$ is a theorem of $\mathcal{T} + \mathcal{T}'$, while $f(p,q,r)$ and $f(p,p,p)$ are distinct irreducibles.

This is not an isolated example. We show now that the same idea applies whenever $\mathcal{T}$ entails a disjunction of equalities without entailing any of the disjuncts.

Proposition 5. *Suppose that for some theory $\mathcal{T}$ and its terms $u_1, v_1, \ldots, u_k, v_k$ the statement*

$$\mathcal{T} \models u_1 \approx v_1 \vee \cdots \vee u_k \approx v_k$$

is true, but none of the statements $\mathcal{T} \models u_i \approx v_i$ is true. Then there exists an equational theory $\mathcal{T}'$ such that $\sigma * \sigma'$ is not a canonizer for $\mathcal{T} + \mathcal{T}'$, for any canonizers σ, σ' of $\mathcal{T}$ and $\mathcal{T}'$.

Proof. Take the signature of $\mathcal{T}'$ to consist of one constant c and one function symbol f of arity $2k$. Axiomatize $\mathcal{T}'$ by k formulas

$$\begin{aligned} f(z,z,x_2,y_2,\ldots,x_k,y_k) &\approx c \\ f(x_1,y_1,z,z,\ldots,x_k,y_k) &\approx c \\ &\cdots \\ f(x_1,y_1,x_2,y_2,\ldots,z,z) &\approx c \end{aligned}$$

It easily follows that $\mathcal{T} + \mathcal{T}' \models f(u_1,v_1,u_2,v_2,\ldots,u_k,v_k) \approx c$. Assuming that there exist canonizers σ, σ' such that $\sigma * \sigma'$ is a canonizer for $\mathcal{T} + \mathcal{T}'$ the corresponding normal forms of the pure $\mathcal{T}'$-term c and the mixed term $f(u_1,v_1,u_2,v_2,\ldots,u_k,v_k)$ must be the same. The mixed term must then be reducible. There is no loss of generality in assuming that the terms u_i, v_i are σ-reduced, so $f(u_1,v_1,u_2,v_2,\ldots,u_k,v_k)$ has only one redex, which is the root position. The result of the alien abstraction of this term is of the form $f(x_1,y_1,x_2,y_2,\ldots,x_k,y_k)$, where x_i, y_i are variables, some of which may be equal (because there may be equals among the terms u_i, v_i). However, we know that $x_i \neq y_i$ for any i (because $u_i \not\approx v_i$ is $\mathcal{T}$-satisfiable). On the other hand, is easy to see that $\mathcal{T}' \models f(x_1,y_1,x_2,y_2,\ldots,x_k,y_k) \approx c$ cannot be true unless $x_i = y_i$ for some i. Consequently, the normal forms of $f(u_1,v_1,u_2,v_2,\ldots,u_k,v_k)$ and c cannot be equal—a contradiction. □

In Theorem 4 below we prove that convexity of the component theories guarantees that the candidate canonizer for their union is indeed a canonizer. This is as much as we can hope for, in view of the examples given in Proposition 5.

For use in the proof of Theorem 4 we need the following modification of the theorem of Tinelli and Harandi about satisfiability in the disjoint union of theories.

Theorem 3. *Let $\mathcal{T} = \mathcal{T}_1 + \cdots + \mathcal{T}_n$, where the theories $\mathcal{T}_i$ are convex, and let ϕ_i be a conjunction of $\mathcal{T}_i$-literals ($i = 1,\ldots,n$). Suppose the set V of variables occurring in all the ϕ_i has at least two elements, and let Δ be the conjunction of all disequations $x \not\approx y$, where $x, y \in V$ and $x \neq y$. If $\phi_i \wedge \Delta$ is $\mathcal{T}_i$-satisfiable for every i, then $\phi_1 \wedge \cdots \wedge \phi_n \wedge \Delta$ is $\mathcal{T}$-satisfiable.*

Remark. The original result ([21], Proposition 3.8) differs from Theorem 3 mainly in that it assumes that the theories $\mathcal{T}_i$ are stably-infinite, rather than convex. For all practical purposes, convexity is

a stronger assumption than stable-infiniteness, as shown recently by Barrett, Dill, and Stump ([5], Theorem 4). Still, there exist convex theories that are not stably-infinite, so Theorem 3 does not directly follow from known results. The proof below is based on ideas in [21] and [5].

Proof. Suppose $V = \{x_1, \ldots, x_m\}$. We prove first that, for every $i \in \{1, \ldots, n\}$, the theory $\mathcal{T}_i' = \mathcal{T}_i \cup \{(\exists \bar{x})\phi_i \wedge \Delta\}$ has an infinite model. (The notation $\bar{x}$ is for the string of variables $x_1, \ldots, x_m$.) Assume the contrary. By Compactness Theorem, there is a finite upper bound k on the set of cardinalities of models of $\mathcal{T}_i'$. Thus, with variables $y_0, \ldots, y_k$ that do not occur in V, we have

$$\mathcal{T}_i, \phi_i \wedge \Delta \models \bigvee_{r \neq s} y_r \approx y_s.$$

Equivalently,

$$\mathcal{T}_i, \phi_i \models \bigvee_{p \neq q} x_p \approx x_q \vee \bigvee_{r \neq s} y_r \approx y_s.$$

Convexity of $\mathcal{T}_i$ implies

$$\mathcal{T}_i, \phi_i \models x_p \approx x_q \text{ for some } p, q$$

or

$$\mathcal{T}_i, \phi_i \models y_r \approx y_s \text{ for some } r, s.$$

The first relation contradicts $\mathcal{T}_i$-satisfiability of $\phi_i \wedge \Delta$. The second even asserts that $\mathcal{T}_i \cup (\exists \bar{x})\phi_i$ can only have a one-element model, which again contradicts $\mathcal{T}_i$-satisfiability of $\phi_i \wedge \Delta$.

Thus, $\mathcal{T}_i'$ has an infinite model, and by the Löwenheim-Skolem Theorem, it has a countably infinite model, say M_i. This M_i is a model for $\mathcal{T}_i$ in which the formula $\phi_i \wedge \Delta$ is satisfiable, via some interpretation that associates distinct elements $a_{i1}, \ldots, a_{ip}$ to variables $x_1, \ldots, x_p$. It is no loss of generality to assume that the underlying sets of models $M_1, \ldots, M_n$ are all the same, and that equalities $a_{1j} = \cdots = a_{mj}$ hold for all $j \in \{1, \ldots, n\}$. (The underlying sets, if different, can be identified via bijections that respect interpretations of the variables x_i.) This common underlying set now becomes a model of $\mathcal{T}$ in which $\phi_1 \wedge \cdots \wedge \phi_n \wedge \Delta$ is satisfiable. (For more details about this "fusion" technique of constructing models of unions of theories, see [2,20].) □

Theorem 4. *Let $\mathcal{T} = \mathcal{T}_1 + \cdots + \mathcal{T}_n$, where each $\mathcal{T}_i$ is a convex theory with a canonizer σ_i. Then $\sigma_1 * \cdots * \sigma_n$ is a canonizer for $\mathcal{T}$.*

Proof. We shall write σ for $\sigma_1 * \cdots * \sigma_n$. Recall that a term t is irreducible precisely when $\sigma(t) = t$. In view of Corollary 4, it suffices to prove that $u \not\approx v$ is $\mathcal{T}$-satisfiable for every two distinct irreducibles u and v. Using Theorem 3, we can translate this $\mathcal{T}$-satisfiability problem to a set of simpler $\mathcal{T}_i$-satisfiability problems. The necessary first step is to transform $u \not\approx v$ to an equisatisfiable conjunction of $\mathcal{T}_i$-formulas, which is commonly done by breaking down mixed terms using variable abstraction repeatedly.

Formally, we let X_0 be the set of variables occurring in u and v, and we let A be the smallest set of terms that contains u and v, and is closed under taking alien subterms. (Thus, the elements of A are

of the form w_π, where w is u or v, and π is the root position of a block of w.) Then we associate a variable $x(t) \notin X_0$ to every $t \in A$, making sure that the map $t \mapsto x(t)$ is injective and order-preserving. Next, with every $t \in A$, we associate the equation

$$E(t): \quad x(t) \approx t[\pi \mapsto x(t_\pi)]^{\pi \in P},$$

where P is the set of alien positions of t. Let us use the shorthand $e(t)$ for the term occurring on the right-hand side of $E(t)$. Each $e(t)$ is a pure $\mathcal{T}_i$-term, for some i. Moreover, since u and v are irreducible, each $e(t)$ is a canonical form for its theory $\mathcal{T}_i$. Note also that the terms $e(t)$ are all distinct.

Let $A = A_1 + \cdots + A_n$ be the partition such that $t \in A_i$ when $e(t)$ is a $\mathcal{T}_i$-term. Let also X_i be the corresponding set of variables $x(t)$. Note that the sets $X_0, X_1, \ldots, X_n$ are disjoint.

Let ϕ_i be the conjunction of equations $E(t)$ where $t \in A_i$. Clearly, ϕ_i is a $\mathcal{T}_i$-formula. We also have

$$\mathcal{T} \models \phi_1 \wedge \cdots \wedge \phi_n \longrightarrow t \approx x(t)$$

for every $t \in A$, by induction on the size of t. As a consequence, proving that $\phi_1 \wedge \cdots \wedge \phi_n \wedge x_u \not\approx x_v$ is $\mathcal{T}$-satisfiable will imply our goal that $u \not\approx v$ is $\mathcal{T}$-satisfiable.

We proceed to prove that $\phi_1 \wedge \cdots \wedge \phi_n \wedge \Delta$ is $\mathcal{T}$-satisfiable, where Δ is the conjunction of disequations $x_s \not\approx x_t$, for all distinct terms $s, t \in A$. By Theorem 3, it suffices to check that $\phi_i \wedge \Delta$ is $\mathcal{T}_i$-satisfiable for each i.

Now, the set of equations occurring in ϕ_i is in solved form for variables in X_i: every $x \in X_i$ occurs once as a left-hand side, and does not occur at all in the right-hand sides. Thus, for any formula ψ, we have that $\phi_i \wedge \psi$ is $\mathcal{T}_i$-satisfiable if and only if the associated formula $\psi' = \psi[x(t) \mapsto e(t)]^{t \in A_i}$ is $\mathcal{T}_i$-satisfiable. We need the instance $\psi = \Delta$ of this observation. It reduces our goal to checking that the formula Δ' is $\mathcal{T}_i$-satisfiable.

The conjuncts of Δ' are disequations each side of which is either a variable in $X - X_i$, or a term $e(t)$ where $t \in A_i$. Thus, every conjunct in Δ' is a disequation of two distinct terms in $T_{\Sigma_i}(X - X_i)$, which are both canonical for $\mathcal{T}_i$. Therefore, by definition of canonizer, each of these disequations is $\mathcal{T}_i$-satisfiable. The convexity of $\mathcal{T}_i$ then implies that their conjunction Δ' is $\mathcal{T}_i$-satisfiable as well. □

8. Non-Existence of Solvers

In this section we show that in general it is not possible to combine Shostak solvers. By Theorem 5 below, most disjoint unions of theories just do not have a solver.

A *general solution* of an equation $u \approx v$ is a set of equations

$$x_1 \approx t_1, \ldots, x_k \approx t_k$$

such that

$$\mathcal{T} \models u \approx v \longleftrightarrow (\exists y_1 \ldots y_l)\ (x_1 \approx t_1 \wedge \cdots \wedge x_k \approx t_k)$$

where: (1) $x_1, \ldots, x_k$ are the variables occurring in $u \approx v$; (2) $y_1, \ldots, y_l$ are the variables occurring in $t_1, \ldots, t_k$; (3) $y_i \neq x_j$ for all i, j. Note that in this situation, $\mathcal{T} \models \theta(u) = \theta(v)$, where θ is the substitution mapping each x_i to t_i.

A *solver* for a theory $\mathcal{T}$ is a function solve that takes an equation $u \approx v$ as argument, and returns a general solution for $u \approx v$ if this equation is $\mathcal{T}$-satisfiable. If $u \approx v$ is $\mathcal{T}$-unsatisfiable, then $\mathsf{solve}(u \approx v)$ returns $\bot$.[3]

In some trivial cases, it is possible to combine solvers. Suppose, for example, that $\mathcal{T}$ is a theory in which all function symbols are "projections" in the sense that $\mathcal{T} \models f(x_1, \ldots, x_n) = x_i$ holds for some i. It is not hard to see that then $\mathcal{T} + \mathcal{T}'$ has a solver for every theory $\mathcal{T}'$ which has a solver. It turns out that these are pretty much all the cases when a combined theory allows a solver.

Let us say that a function symbol f (of any non-zero arity) of $\mathcal{T}$ is *non-collapsing* when $f(x, \ldots, x) \not\approx x$ is $\mathcal{T}$-satisfiable.

Theorem 5. *Suppose $\mathcal{T}_1$ and $\mathcal{T}_2$ are consistent stably-infinite theories with non-collapsing function symbols, and suppose σ_1, σ_2 are canonizers of these theories. If $\sigma_1 * \sigma_2$ is a canonizer for $\mathcal{T} = \mathcal{T}_1 + \mathcal{T}_2$, then $\mathcal{T}$ does not have a solver.*

Proof. Consider the equation

$$E: \quad f(x, \ldots, x) \approx g(x, \ldots, x)$$

where f and g are non-collapsing symbols of $\mathcal{T}_1$ and $\mathcal{T}_2$ respectively. Note first that the theory $\mathcal{T}_1 + \mathcal{T}_2$ is consistent: each component theory has a countably infinite model by stable infiniteness, and the two models can be "fused" as desribed in [2] to produce a model for $\mathcal{T}$ (cf. proof of Theorem 3 above). We need to check that both E and $\neg E$ are $\mathcal{T}$-satisfiable. Indeed, since $\mathcal{T}_1$ is stably infinite, it has a countably infinite model M_1 which contains distinct elements a_1 and b_1 such that $f^{M_1}(a_1, \ldots, a_1) = b_1$. Similarly, there is a countably infinite model M_2 of $\mathcal{T}_2$ containing distinct elements a_2 and b_2 such that $g^{M_2}(a_2, \ldots, a_2) = b_2$. Every bijection between the carrier sets of these models produces a "fusion" model for $\mathcal{T}_1 + \mathcal{T}_2$ [2]. Choosing the bijection so that a_1 corresponds to a_2 and b_1 corresponds to b_2 will result in a model satisfying E. Another choice, where a_1 corresponds to a_2 but b_1 does not correspond to b_2 will give a model satisfying $\neg E$.

Arguing by contradiction, assume there exists a solver for $\mathcal{T}$. Since E is satisfiable, $\mathsf{solve}(E)$ is an equation of the form $x \approx w$, where x does not occur in w. It follows that

$$\mathcal{T} \models f(w, \ldots, w) \approx g(w, \ldots, w) \tag{3}$$

and, since $\sigma_1 * \sigma_2$ is a canonizer, the normal forms of $f(w, \ldots, w)$ and $g(w, \ldots, w)$ must be the same. We proceed to show that their normal forms must also be distinct.

We may assume without loss of generality that w is irreducible. Since $\neg E$ is satisfiable, it follows from (3) that w cannot be x (or any other variable). For definiteness, suppose the top symbol of w is in $\mathcal{T}_1$.

The only possible redex of the term $g(w, \ldots, w)$ is ϵ. Since g is a non-collapsing symbol, $\sigma_2(g(x, \ldots, x))$ is not a variable, but some proper $\mathcal{T}_2$-term. Thus, reduction will not change the block height[4] of

[3] The power of effective solvers is in their ability to reduce the decidability problem for Horn clauses over a given theory $\mathcal{T}$ to the word problem for $\mathcal{T}$; see [5] and [9].

[4] The *block height* of a term t is the maximum number of blocks one can visit on a path from the root to a leaf of the tree of t.

$g(w, \ldots, w)$, which is one greater than the block height of w, since g and the top symbol of w belong to different signatures.

On the other hand, the block height of $f(w, \ldots, w)$ clearly equals that of w and cannot increase when $f(w, \ldots, w)$ is reduced. (By definition of reduction, block height of mixed terms cannot increase at any reduction step.) Thus, $f(w, \ldots, w)$ and $g(w, \ldots, w)$ have different normal forms. □

9. Conclusion and related work

Along with the combination algorithm of Nelson and Oppen [13], the method suggested by Shostak [19] has been a cornerstone for implementation of automated verification tools based on combining decision procedures. In a recent survey [17], Shankar discusses the promise and success of such tools, stressing also the need for stronger theoretical support. Clarifying theoretical foundations of the area has become a subject of intensive research; the list [3,9,11,18,12,7] is a sample from the spate of recent publications. Much of this effort, including the present paper, is devoted to the demystification of the Shostak method. Our contribution is in providing answers to two basic questions that have not as yet been adequately addressed.

With Theorem 4 we confirm the common view that canonizers for disjoints unions of theories can be obtained by a straightforward combination of canonizers for the component theories. Our analysis reveals also that this result only holds with some additional assumptions on the theories involved, and that convexity of theories is a sufficient condition.

It is not clear whether the solvability of the word problem for a theory implies the existence of an effective canonizer for it. If so, our Theorem 4 could be viewed as a generalization of Pigozzi's theorem [15] which states that the word problem is solvable for disjoint unions of *equational* theories with solvable word problems. Pigozzi's result has recently been revisited and generalized in a different direction by Baader and Tinelli [2]. In fact, their version of the algorithm for combining solutions of the word problem for disjoint equational theories remains correct even for some sets of non-equational input theories. It appears that this algorithm correctly works for any set of theories whose canonizers are combinable, and that this could be proved by exploiting the characterization of combinability given in our Theorem 2.

Combination of canonizers is a basic technique that provides grounds for equational reasoning about terms in unions of theories, much like normal forms in various colimits of algebraic structures do. We expect therefore Theorem 4 to be of wider interest and applicability. Its usefulness is demonstrated by an application in [10], and also in our proof of Theorem 5.

Theorem 5 itself confirms another observation, made only recently [5,18], namely that there is no general way of producing a solver for the disjoint union of theories from solvers of the component theories. Acknowledging this fact, and thus abandoning the idea of producing the combined solver altogether, the designers of the prover ICS make decision procedures of Shostak theories cooperate in a Nelson-Oppen framework, reducing the role of Shostak solvers to efficient generation of new equalities [18].

On the other hand, Theorem 5 implies that a direct combination of solvers is not possible for theories of practical interest, and this seems to contradict the common wisdom, as well as practice, where some tools (e.g. CVC, as described in [3]) apparently combine solvers of several Shostak theories into a global solver. This conundrum needs to be resolved, but it would be premature to claim

that Theorem 5 destroys the possibility of global solvers. Perhaps such solvers exist in some modified setting that has not been fully explained yet. With this additional motivation, we would join Tinelli and Ringeissen [20] in their call to investigate combining decision procedures for *multisorted* theories.

Acknowledgments

We thank Andrew Tolmach for inciting this research and for comments. We thank Cesare Tinelli for numerous corrections and suggestions. We also thank Nikolaj Bjørner, Kosta Došen, John Matthews, Natarajan Shankar, Tim Sheard, and the anonymous CADE-19 referees for their valuable feedback.

References

[1] F. Baader, T. Nipkow, Term rewriting and all that, Cambridge University Press, United Kingdom, 1998.
[2] F. Baader, C. Tinelli, Deciding the word problem in the union of equational theories, Information and Computation 178 (2002) 346–390.
[3] C. Barrett, Checking Validity of Quantifier-free formulas in Combinations of First-Order Theories, PhD thesis, Stanford University, 2002.
[4] C. Barrett, D. Dill, J. Levitt, Validity checking for combinations of theories with equality, in: M. Srivas, A. Camilleri (Eds.), Formal Methods In Computer-Aided Design, Lecture Notes in Computer Science, vol. 1166, Springer, 1996, pp. 187–201.
[5] C.W. Barrett, D.L. Dill, A. Stump, A generalization of Shostak's method for combining decision procedures, in: Frontiers of Combining Systems (FROCOS), Lecture Notes in Artificial Intelligence, vol. 2309, Springer, 2002, pp. 132–147.
[6] N. Bjørner, et al., Deductive-algorithmic verification of reactive and real-time systems, in: R. Alur, T.A. Henzinger (Eds.), Computer-Aided Verification (CAV), Lecture Notes in Computer Science, vol. 1102, Springer, 1996, pp. 415–418.
[7] S. Conchon, S. Krstić, Strategies for combining decision procedures, in: P. Narendran, M. Rusinowitch (Eds.), Tools and Algorithms for the Construction and Analysis of Systems (TACAS), Lecture Notes in Computer Science, vol. 2619, Springer, 2003, pp. 537–553.
[8] D. Cyrluk, P. Lincoln, N. Shankar, On Shostak's decision procedure for combinations of theories, in: M.A. McRobbie, J.K. Slaney (Eds.), Automated Deduction (CADE-13), Lecture Notes in Artificial Intelligence, 1104, Springer, 1996, pp. 463–477.
[9] H. Ganzinger, Shostak light, in: A. Voronkov (Ed.), Automated Deduction (CADE-18), Lecture Notes in Artificial Intelligence, vol. 2392, Springer, 2002, pp. 332–347.
[10] H. Ganzinger, T.h. Hillenbrand, U. Waldmann, Superposition modulo a Shostak theory, in: F. Baader (Ed.), Automated Deduction (CADE-19), LNAI, vol. 2741, Springer-Verlag, 2003, pp. 182–196.
[11] D. Kapur, A rewrite rule based framework for combining decision procedures, in: Frontiers of Combining Systems (FROCOS), Lecture Notes in Artificial Intelligence, vol. 2309, Springer, 2002, pp. 87–103.
[12] Z. Manna, C.G. Zarba, Combining decision procedures, in: Formal Methods at the Cross Roads: From Panacea to Foundational Support, Lecture Notes in Computer Science, vol. 2757, Springer, 2003, pp. 381–422.
[13] G. Nelson, D.C. Oppen, Simplification by cooperating decision procedures, ACM Transactions on Programming Languages and Systems 1 (2) (1979) 245–257.
[14] S. Owre, J. Rushby, N. Shankar, F. von Henke, Formal verification for fault-tolerant architectures: prolegomena to the design of PVS, IEEE Transactions on Software Engineering 21 (2) (1995) 107–125.
[15] D. Pigozzi, The join of equational thories, Colloquium Mathematicum 30 (1974) 15–25.

[16] H. Rueß, N. Shankar, Deconstructing Shostak, In Proceedings of the 16th Annual IEEE Symposium on Logic in Computer Science (LICS '01), IEEE Computer Society, 2001, pp. 19–28.
[17] N. Shankar, Little engines of proof, in: L. Eriksson, P. Lindsay (Eds.), FME 2002: Formal Methods—Getting IT Right, Springer, Copenhagen, 2002, pp. 1–20.
[18] N. Shankar, H. Rueß, Combining Shostak theories, in: S. Tison (Ed.), Rewriting Techniques and Applications (RTA), Lecture Notes in Computer Science, vol. 2378, Springer, 2002, pp. 1–19.
[19] R.E. Shostak, Deciding combinations of theories, Journal of the ACM 31 (1) (1984) 1–12.
[20] C. Tinelli, C. Ringeissen, Unions of non-disjoint theories and combinations of satisfiability procedures, Theoretical Computer Science 290 (2003) 291–353.
[21] C. Tinelli, M.T. Harandi, A new correctness proof of the Nelson–Oppen combination procedure, in: F. Baader, K.U. Schulz (Eds.), Frontiers of Combining Systems: Proceedings of the 1st International Workshop, Kluwer, 1996, pp. 103–120.